Is the Large Simple Trial Design Used for Comparative, Post-Approval Safety Research?
A Review of a Clinical Trials Registry and the Published Literature

Robert F. Reynolds,[1,2] *Joanna A. Lem,*[1] *Nicolle M. Gatto*[1,2] and *Sybil M. Eng*[3]

1 Department of Epidemiology, Pfizer, Inc., New York, NY, USA
2 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
3 Global Health Insights, LLC, Montclair, NJ, USA

Contents

Abstract

Post-approval, observational drug safety studies face well known difficulties in controlling for confounding, particularly confounding by indication for drug use. A study design that addresses confounding by indication is the large simple trial (LST). LSTs are characterized by large sample sizes, often in the thousands; broad entry criteria consistent with the approved medication label; randomization based on equipoise, i.e. neither physician nor patient believes that one treatment option is superior; minimal, streamlined data collection requirements; objectively-measured endpoints (e.g. death, hospitalization); and follow-up that minimizes interventions or interference with normal clinical practice. In theory then, the LST is a preferred study design for drug and vaccine safety research because it controls for biases inherent to

observational research while still providing results that are generalizable to 'real-world' use.

To evaluate whether LSTs are used for comparative safety evaluation and if the design is, in fact, advantageous compared with other designs, we conducted a review of the published literature (1949 through 31 December 2010) and the ClinicalTrials.gov registry (2000 through 31 December 2010). Thirteen ongoing or completed safety LSTs were identified. The design has rarely been used in comparative drug safety research, which is due to the operational, financial and scientific hurdles of implementing the design. The studies that have been completed addressed important clinical questions and, in some cases, led to re-evaluation of medical practice. We conclude the design has demonstrated utility for comparative safety research of medicines and vaccines if the necessary scientific and operational conditions for its use are met.

Observational designs and methods are important for evaluating the safety of medicines post-approval. They may be the only means to study large populations under routine clinical conditions or to evaluate a medicine's association with rare events or long-latency outcomes. As such, the use of epidemiological designs and methods to assess the safety of a new medicine after it enters the market is a logical complement to pre-approval safety data collected from randomized controlled trials (RCTs). In the past 10 years, requirements for epidemiological studies as a condition of regulatory approval for new medicines have increased.[1,2] It is now generally recognized that an evaluation of a drug's overall safety profile is strengthened by data on its use among real-world populations.[3-7]

However, despite these advances, the role of findings from observational studies for regulatory and clinical decision-making is frequently contested. An important reason is that epidemiological studies of medication exposures and their effects have difficulty measuring and controlling for confounding by indication for drug use (and/or severity of disease).[8-10] As with any form of confounding, an investigator can, in theory, control for confounding by indication if its causes or their surrogates can be identified and reliably measured. In practice this is not easily done. Concerns about uncontrolled confounding by indication are central to debates about the interpretation of observational study findings and have complicated the assessment of many potential medicine-adverse outcome associations.[11-22]

Nonetheless, the options for investigating real-world safety, or the safety of a medicine as it is actually prescribed by physicians and used by patients once on the market, are limited.[23] Post-approval phase IV RCTs are ethically and logistically difficult to conduct, are not appropriate for addressing many safety issues (e.g. rare or longer-term outcomes) and are unlikely to provide information on the safety of medicines in real-world populations. Thus, studies that are observational and follow patients with minimal interference are often the only means by which to answer clinically meaningful questions about a marketed medicine's safety and to study patients in settings that generalize to real-world medicine use.

The ideal design for a post-approval safety study is one that minimizes the potential for bias yet is still relevant to real-world clinical practice. A design that, in principle, merges the ideal characteristics of the RCT (randomization) with those of an observational epidemiology study (follow-up with minimal intervention) is the large simple trial (LST).[24-26] LSTs are characterized by large sample sizes, often in the thousands; broad entry criteria consistent with the approved medication label; randomization based on equipoise, i.e. neither physician or patient believes that one treatment option is superior; minimal data requirements such as a questionnaire or case report form usually collecting data in only a few pages with questions limited to key variables that are typically collected at routine clinical care visits; objectively-measured endpoints (e.g. death,

hospitalization); follow-up that minimizes interventions or interference with normal clinical practice; follow-up of all patients regardless of whether they discontinue randomized medication; and intent-to-treat (ITT) analysis examining the entire population of randomized subjects according to the treatment group to which they were initially randomized.

Although the LST design shares a key design component, randomization, with an RCT, it is distinguished by its intent to minimize interference with usual medical care. Notably, since both designs employ randomization to allocate patients to treatment groups they are defined as interventional studies. In the RCT, the intent is restriction and control to create experimental conditions at baseline and over the course of the study. In contrast, in the LST the intent is to create balance of baseline characteristics but to then follow patient outcomes using observational methods. Practically, this means that endpoint definition, physician and patient recruitment, drug delivery, data collection, allowance for treatment discontinuation, concomitant drug use, and patient and site monitoring in an LST are operationally different from that of an RCT. Table I highlights the key aspects that differentiate these designs.

Calls for the use of the LST design have, until recently, primarily focused on the need for studies of clinically important therapeutic or preventative effects of interventions to inform clinical and health policy decisions.[26,27] LSTs have been employed to study a range of real-world benefits, from interventions for the treatment and prevention of cardiovascular outcomes[28-32] to a comparison of antiretroviral treatment strategies for HIV-positive patients.[33] In the 1990s, two research groups described its use for safety, each using a single LST they had completed as a case study to infer general lessons for future studies.[34,35] More recently, because of its unique design characteristics, the LST has been put forward in regulatory guidance from the US FDA[36] and the Institute of Medicine's (IOM) Report on Drug Safety[37] as having the potential to further characterize a drug's post-approval safety profile. To evaluate whether the design is in fact advantageous in practice, we conducted a review of the published literature and the ClinicalTrials.gov registry. The aim was to identify studies that used the design to study safety outcomes and to determine if those studies had successfully answered their original hypotheses. The intent of our analyses was to draw conclusions about the utility of the LST

Table I. Design characteristics of a large simple trial (LST) compared with those of a randomized controlled trial (RCT)

Design characteristic	LST	RCT
Randomization	Yes	Yes
Sample size	Larger (thousands)	Smaller (hundreds)
Inclusion criteria	Broad (e.g. per approved drug label)	Narrow (e.g. excludes patients with co-morbid conditions, using multiple medications, pregnant women, elderly)
Questionnaire/case report form/interview	Minimal, brief	Complex, lengthy
Endpoints	'Hard' endpoints (mortality, hospitalization or life-threatening events)	All
Required patient visits and procedures	Few, if any; follows normal practice schedule and assessments	Yes, frequent; visits and tests far greater than expected in clinical practice
Primary source of investigators or enrolling physicians	General practitioners/community-based	Clinical research/academic centres
Site monitoring	Minimal	Frequent
Patient followed after randomized treatment discontinued	Yes	No, or for limited duration post-discontinuation (e.g. 30 days)
Primary analytic method	ITT	ITT

ITT = intent-to-treat.

Reynolds et al.

design, and the conditions under which it is most appropriate, for comparative safety research.

The specific objectives of this review were to (i) identify all ongoing or completed LSTs with a primary safety endpoint described in the public domain; (ii) analyse and summarize the key design aspects and results of these studies, including whether the study successfully addressed the original research question(s); and (iii) describe any design characteristics that were noted by study authors to complicate the conduct of the study and interpretation of results.

1. Data Sources and Search Strategy

We conducted a systematic review of the published medical literature using PubMed and the ClinicalTrials.gov registry. The process used to identify studies is described in the following sections (a more detailed description of the criteria used is available in Appendix I, Supplemental Digital Content 1, http://links.adisonline.com/DSZ/A52). In addition to the systematic search of the literature and ClinicalTrials.gov registry, we reviewed bibliographies of LST review articles and book chapters, and conducted an informal survey of international experts in drug safety. Many of the responders to the informal survey are past or current investigators of the safety LSTs identified in this review (a list of the process and respondents is available in Appendix II, online SDC 1).

1.1 Identifying Large Simple Trials (LSTs)

For the purpose of this review, we defined 'LST' as any randomized study with simplified study procedures permitting comparative assessment of medicines under real-world or routine clinical conditions. Consistent with the focus on studies of real-world, post-approval safety, we sought to identify studies that principally used observational follow-up post-randomization, i.e. had few requirements post-randomization that differed from a physician's prescribing practice or patient's usual care. Formally and colloquially, these studies are referred to by many names, including LSTs/studies, large simplified trials/studies, large streamlined trials/studies, naturalistic trials/studies,

practical clinical trials and pragmatic clinical trials. Thus, our searches used such terms as 'simple', 'simplified', 'streamline', 'pragmatic', 'practical', 'naturalistic', 'trial', 'study' and 'randomization' to capture any randomized study with simplified procedures that allowed routine care conditions. The prospective, randomized, open-label, blinded endpoint evaluation (PROBE) term was not included in the search string because of its inconsistent use in the literature. Furthermore, while many LSTs meet the criteria of a PROBE design most PROBE studies are not LSTs.

1.2 Identifying Studies with a Primary Safety Endpoint

The goal of this review was to identify whether the LST design had been used for the primary purpose of testing differences in safety outcomes. Therefore, the review excluded simple trials in which the primary hypothesis/endpoint was an efficacy or effectiveness endpoint. We used search terms such as 'safety' and 'risk' to limit the retrieved studies to those that included a safety endpoint or evaluated the comparative safety profile of the health interventions being studied in the trial. Clearly, randomized efficacy and effectiveness studies collect data on safety, as required by regulations, and may include a secondary endpoint of safety. However, in our review of abstracts and published manuscripts we sought to distinguish studies designed and powered to test a pre-specified primary safety hypothesis from those that included one or more secondary safety hypotheses or no specified safety hypothesis.

2. Systematic Literature Search

2.1 PubMed Search Strategy

PubMed is a service of the US National Library of Medicine that includes over 20 million citations from MEDLINE, life science journals and online books. We searched this database, including all citations posted through 31 December 2010. RCTs, drug toxicity and adverse events were searched as MESH terms and subheadings (where relevant). Terms specific to LSTs were included using the 'TW' search field tag, which

searches the entire record including the article title, abstract and all indexing terms associated with the article. The search string was as follows: ('Randomized Controlled Trial' [Publication Type]) AND (drug toxicity [MeSH] OR adverse effects [MeSH]) AND ('large scale' [TW] OR simple [TW] OR simplistic [TW] OR pragmatic [TW] OR practical [TW] OR streamline [TW] OR naturalistic [TW]).

2.2 ClinicalTrials.gov Search Strategy

ClinicalTrials.gov (www.ClinicalTrials.gov) is a registry of federally and privately supported clinical trials conducted in the US and around the world. Clinical trials in this registry are defined broadly as any research study in human volunteers to answer specific health questions; thus, the registry includes interventional and observational studies. The first version of the system was publicly available in February 2000.[38] Organizations that sponsor and implement clinical studies are responsible for submitting accurate and timely information about their studies to ClinicalTrials.gov, which is then reviewed by the National Library of Medicine (NLM). The database is updated daily. At the time of the registry search, ClinicalTrials.gov contained more than 100 000 trials sponsored by the National Institutes of Health, other federal agencies, and private industry. Studies listed in the database were conducted in all 50 states of the US and 174 countries.

We searched the ClinicalTrials.gov registry, including all studies posted February 2000 through 31 December 2010. The registry was searched in a similar fashion to the medical literature, albeit with variations because of its specific search requirements. The search string was as follows: ('simple' OR 'simplified' OR 'streamline' OR 'pragmatic' OR 'practical' OR 'naturalistic') AND 'randomized' AND 'safety study'. In addition, the search was restricted to 'Interventional Studies' and 'Phase IV' using the registry selection options.

3. Criteria for Evaluating Abstract Summaries and Published Papers

After conducting the searches, identification of studies for inclusion in the safety LST analysis

was conducted in two stages. Each stage had specific inclusion and exclusion criteria. In Stage 1, abstracts and summaries were reviewed and studies were identified for full paper review. In Stage 2, full papers were reviewed and studies were identified for inclusion in the analysis.

3.1 Stage 1: Selecting Abstracts and Clinicaltrials.gov Summaries for Full Paper Review

Each abstract and study summary was reviewed independently by two co-authors. The study was included in Stage 2 if at least one reviewer identified the abstract or study summary as meeting the specified inclusion/exclusion criteria or if there was insufficient information to evaluate the abstract or study summary. The inclusion and exclusion criteria are described in Appendix I (online SDC 1) and summarized below.

3.1.1 Inclusion Criteria

1. Investigators describe a specific primary safety endpoint, set of safety endpoints or a composite safety or risk score.
2. Study randomizes patients to a specific health intervention (treatment, vaccine, device, treatment strategy) or an appropriate usual-care comparison group.
3. Phase IV study.
4. Conducted at more than one site and includes at least 100 patients.

3.1.2 Exclusion Criteria

1. Studies comparing the (general) adverse event profiles of two or more interventions without a specific endpoint of interest.
2. Studies whose primary endpoint is a Quality of Life or general well-being measure.
3. Health education, surgical procedure or dosing studies.
4. Phase I–III studies.
5. Pilots, feasibility assessments, or secondary analyses and extension studies of RCTs.

3.2 Stage 2: Selecting Published Papers for Inclusion in Safety LST Analysis

In Stage 2 of the review, the published articles for the abstracts and summaries identified in

Reynolds et al.

Stage 1 were reviewed independently by two co-authors. If both reviewers identified the study as meeting the specified inclusion/exclusion criteria, the study was included in the analysis. If only one author identified the study as meeting the specified inclusion/exclusion criteria, the authors discussed the study to reach consensus.

We first verified that the study met the inclusion and exclusion criteria set in Stage 1. If so, we reviewed the publication for evidence that the investigators intended to design and conduct a study reflective of real-world practice. Typically, this would be accomplished by studying the indicated population; using study procedures that do not deviate from normal practice, or do so minimally; and using observational follow-up methods to assess outcomes. The inclusion criteria are summarized below and are described in detail in Appendix I (online SDC 1).

3.2.1 Inclusion Criteria

1. Eligibility per approved product label or Summary of Product Characteristics so that patients reflect those seen in a typical practice setting.
2. Measurements and laboratory tests in the study are generally considered standard clinical practice for the treatment under evaluation and/or for the comparison group.
3. Physician visits do not significantly exceed those expected for the indicated population.
4. Data collected that are not directly related to the outcome assessment are limited.
5. Outcome assessment is primarily dependent on a combination of patient or physician report, simple home or outpatient tests/measurements that are consistent with standard practice and home care, or physician and medical/hospital records.

Each study document (published paper or ClinicalTrials.gov summary) selected for inclusion was then reviewed to extract the following data: study name, year of publication or year posted on ClinicalTrials.gov, study drug (generic name), comparator, sample size and the verbatim description of the primary endpoint(s). We then extracted key design information for each study to demonstrate the extent to which the study mimicked routine care and used observational follow-

up. For completeness, we also extracted information about whether the study performed intent-to-treat (ITT) analysis and/or a time on treatment (exposed person-time) analysis, and summarized the study's results, including risk estimates and 95% confidence intervals (CIs) when these were reported.

4. LSTs with a Primary Safety Endpoint

Thirteen LSTs with primary safety endpoints were identified.[39-53] Ten studies were discovered by the systematic search of the published medical literature (n = 7)[41-43,45-47,51,52] and the ClinicalTrials.gov registry (n = 4),[44,49,50,53] with one study identified in both searches,[42-44] two through prior knowledge of the authors,[39,40] and one through the survey of drug safety experts.[48]

The stages of the review process for the PubMed and ClinicalTrials.gov searches are depicted in figure 1.

The PubMed search of literature from 1949 through 2010 retrieved 1323 abstracts. Of these, 1258 were excluded and 65 were determined to meet the criteria for full-paper review. The most common reasons for excluding abstracts were safety was not a primary endpoint in the study; the abstract was selected because the search term referred to something other than study design or procedures; the study did not assess drug treatment effects (e.g. numerous short-term, single centre 'simple' surgical procedure studies); the study was a phase I healthy subject study, phase II dosing study or phase III extension study and thus did not use approved, marketed medicines; or the study design was not randomized. In addition, the search retrieved many abstracts describing phase IV RCT or LST designs where the primary purpose of the study was to assess an efficacy or effectiveness endpoint but where data on the general safety profile were also collected. Of the 65 full papers reviewed, 58 studies were excluded because the study was a methodological study (n = 2); phase I–III study (n = 5); RCT with a primary efficacy endpoint (n = 34); RCT extension follow-up study (n = 4); LST with a primary effectiveness endpoint (n = 2); RCT with a primary safety endpoint (n = 9); or LST evaluating
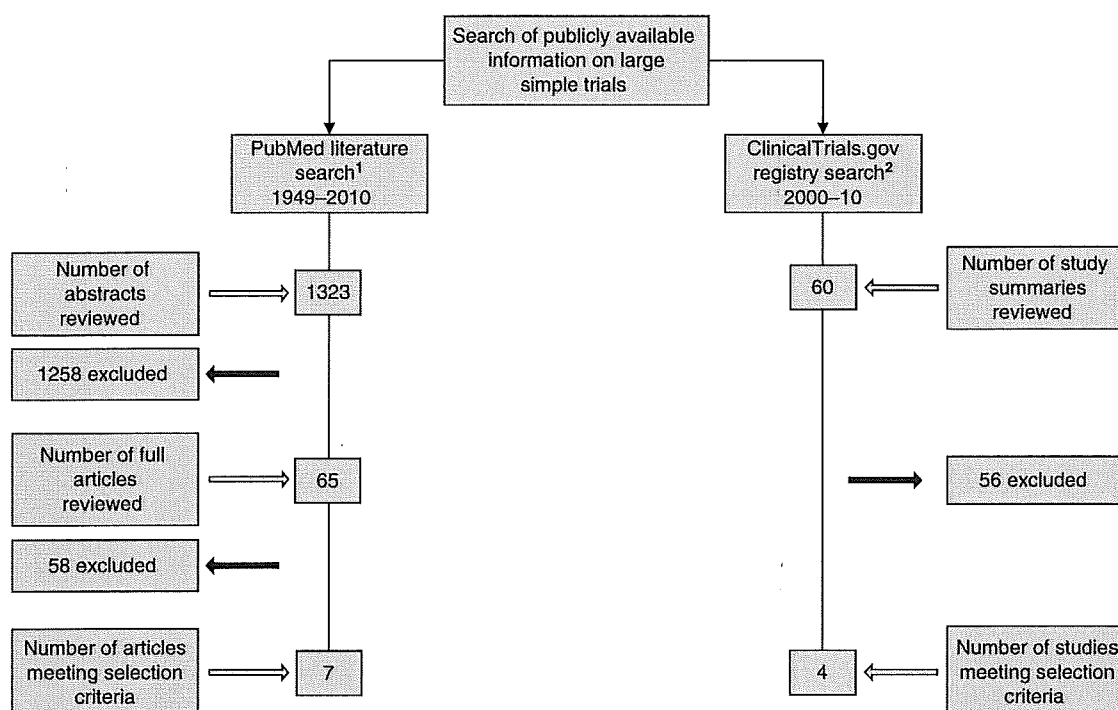
Fig. 1. Results of the medical literature and ClinicalTrials.gov registry search and review process.

safety profile without a specific endpoint (n = 2). Three excluded studies met some of the criteria used to define a safety LST. Two studies were LSTs and evaluated safety but did not specify a primary endpoint.[54,55] The third study[56] evaluated a safety endpoint and followed patients annually for 7 years but required extensive laboratory, psychometric and neurological assessments at baseline and subsequent visits and, thus, did not study subjects under routine clinical care conditions. Overall, seven studies were identified from the published medical literature.

The clinicaltrials.gov search returned 60 study summaries. Four studies met the study selection criteria: the VOLUME Study (ClinicalTrials.gov ID NCT00359801) [see table II for definition of study acronyms], the ZODIAC Study (Clinical-Trials.gov ID NCT00418171), the SCOT Large Simple Safety Trial (ClinicalTrials.gov ID NCT00447759), and the GiSAS pragmatic RCT (ClinicalTrials.gov ID NCT01052389).[44,49,50,53] The ZODIAC Study had already been identified in the medical literature search. Of the 60 studies identified, 56 were excluded because the identi-

fied search term referred to something other than the study design or procedures (e.g. 'simple blind', 'simple verbal scale', 'simplified treatment') [n = 34] or did not include safety as the primary endpoint (n = 22).

Table II. Study acronyms

COSMIC	Comparative Outcomes Study of Metformin Intervention Versus Conventional Approach Study
GiSAS	Aripiprazolo, Olanzapina e Aloperidolo nel trattamento a lungo termine della schizofrenia (aripiprazole, olanazpine and haloperidol in the long-term treatment of schizophrenia)
HYPREN	Hypertension under Prazosin and Enalapril Study
PAIN	Paracetamol, Aspirin and Ibuprofen New Tolerability Study
SCOT/SCOTLSSS	Scottish Large Simple Safety Study
SMART	Salmeterol Multicenter Asthma Research Trial
VOLUME	Exubera Large Simple Trial to Evaluate Long-Term Pulmonary and Cardiovascular Safety
ZODIAC	Ziprasidone Observational Study of Cardiac Outcomes

Two studies[39,40] known to the authors were not identified in the published medical literature search or in the ClinicalTrials.gov registry, and only one study[48] of four potential LSTs[48,57-59] identified by the survey of drug safety experts met the study selection criteria. None of these studies were indexed using the broad search terms we used to describe LSTs. Upon further review, we discovered that the terms to which they were indexed were very general. In order to have retrieved these studies we would have had to use a search strategy retrieving more than 60000 abstracts for review. Finally, all four studies were conducted or began enrolment prior to initiation of the ClinicalTrials.gov registry, therefore we did not expect them to be catalogued there.

In summary, a total of 13 LSTs were identified for analysis through the published medical literature, ClinicalTrials.gov registry, survey of drug experts and prior knowledge of the authors.

5. Comparison of Design Characteristics of the LSTs

The design elements of the 13 studies are summarized in table III.

Upon review of these studies, we found they spanned a period of about 20 years, were conducted in multiple disease areas, and evaluated drugs and vaccines or treatment strategies (treatment duration, type, or dose and number of vaccine injections). Two trials, SCOT and GiSAS, are ongoing.[50,53] Three trials were conducted to examine the safety of prescription and over-the-counter NSAIDs,[40,41,50] two among patients being treated for diabetes mellitus[46,49] and three to assess the safety of atypical antipsychotics for patients with schizophrenia.[43,45,53] Only one study used placebo added to usual care as the comparator.[48] Most of the trials (n = 9) enrolled patients from general or community practice sites,[40-44,46-48,50,53] although four studies were conducted in hospitals.[39,45,51,52] Two studies tested the safety of health interventions among paediatric patients.[40,47]

Nine studies recruited and enrolled patients from one country only,[39-41,45-48,52,53] whereas four studies,[43,49-51] which had large actual or intended sample sizes, included sites from multiple countries. The actual or intended sample sizes of all studies ranged from just over 100 patients to more than 80000 randomized patients. One study, VOLUME,[49] intended to enrol sites from more than 20 countries but did not reach its target sample size because of the market withdrawal of one of the study drugs, inhaled insulin, for commercial reasons. Despite having enrolled more than 25000 patients, the SMART trial also terminated early. The primary outcome event rate was lower than expected and revised estimates of the sample size were approximately 60000 patients.[48] A direct relationship between the sample size of the trial and the number of countries participating was not evident.

The primary safety endpoint for each study ranged from outcomes that required scale or instrument measurement during the course of the study, such as a persistent decline in pulmonary function (spirometry)[49] and tardive dyskinesia (symptom scales)[45] to 'hard' events such as death.[43,48,50] The majority of studies used broad inclusion criteria, had few procedures required and allowed flexible dosing. Inclusion in the studies was primarily determined by clinical opinion, i.e. the physician's interpretation of the drug's approved labelling with minimal or no additional exclusion criteria. Except for the COSMIC,[46] Pediatric Vaccine,[47] and Latent TB[51] studies, the primary analysis for the LSTs was ITT. These analyses included patients in the analysable population even if they discontinued or switched their assigned treatment. Although information was lacking from some studies, we found that in most studies patients were followed for periods that exceeded 30 days after last trial visit. This is in contrast to the approach typically used in many clinical trials where outcome follow-up ends 30 days after the last visit.

6. Results and Conclusions of the Completed LSTs

Each study's rationale, main objective, primary results and conclusions are presented in table IV. The data collection method(s) used to obtain information about study endpoints, the duration of

Table III. Large simple trials (LSTs) with a primary safety endpoint by year of publication or ClinicalTrials.gov posting date

Study design elements							LST characteristics							Analytic method	
study (y)[a,b]	year conducted	study drug	comparison	sample size: actual (intended)	participating countries (no. of sites and/or investigators)	primary endpoint(s)	open-label	general practice	broad inclusion criteria[c]	concomitant medications allowed	flexible dosing	minimal procedures required[d]	infrequent or no visits	per intent-to-treat[e]	per time on treatment[e]
HYPREN (1991)[39]	NR	Enalapril	Prazosin	1210	Germany (101 hospitals)	Incidence of symptomatic hypotension on first d of treatment	✓	–	–	✓	–	✓	✓	✓	–
Pediatric Ibuprofen Study (1995)[40]	1991–3	Ibuprofen	Paracetamol (acetaminophen)	84192	US (1735 paediatricians and GPs)	Hospitalization for acute GI bleeding, acute renal failure, anaphylaxis or Reye's syndrome	–	✓	✓	✓	–	✓	✓	✓	–
PAIN (1999)[41]	1997–8	Ibuprofen	Paracetamol and aspirin (acetylsalicylic acid) prescribed at OTC dosages	8677	France (1108 GPs)	Proportion of patients with at least one significant adverse event	–	✓	✓	✓	✓	✓	✓	✓	✓[g]
ZODIAC (2002)[42-44],i,h	2002–6	Ziprasidone	Olanzapine	18240	Sweden, Brazil, USA, Argentina, Chile, Peru, Taiwan, Thailand, Korea, Malaysia, Hong Kong, Singapore, Uruguay, Romania, Slovakia, Hungary, Poland, Mexico (749 centres)	Non-suicide mortality within 12mo	✓	✓	✓	✓	✓	✓	✓	✓	✓[g]
Olanzapine and Tardive Dyskinesia Study (2004)[45],i	NR	Olanzapine	Typical antipsychotics	190	Brazil (3 hospitals)	TD	✓	–	✓	✓	✓	✓	✓	✓	–

Continued next page

PX187-0009

Reynolds et al.

Table III. Contd

Study design elements							LST characteristics							Analytic method	
study (y)[a,b]	year conducted	study drug	comparison	sample size: actual (intended)	participating countries (no. of sites and/or investigators)	primary endpoint(s)	open-label	general practice	broad inclusion criteria[c]	concomitant medications allowed	flexible dosing	minimal procedures required[d]	infrequent or no visits	per intent-to-treat[e]	per time on treatment
COSMIC (2005)[46]i	NR	Metformin	Usual diabetes mellitus care	9000	US (Not reported)	Incidence of lactic acidosis or other serious adverse events	✓	✓	✓	✓	✓	✓	✓	–	✓
Pediatric Vaccination Strategies Study (2006)[47]	NR	HEXAVAC vaccine	PENTACOQ or HBVAXPRO	7151	France (389 paediatricians)	Incidence of fever of 40°C or higher (rectal)	✓	✓	✓	✓	NA	✓	✓	✓g	✓
SMART (2006)[48]	1996–2003	Salmeterol via metered-dose inhaler+usual care	Placebo via metered-dose inhaler+usual care	26355 (60000)	US (1316 investigators from 6163 sites)	Respiratory-related mortality and respiratory-related, life-threatening experiences	–	✓	✓	✓	–	✓	✓	✓	–
VOLUME (2006)[49]h	2006–9	Exubera	Usual diabetes care	1985 (5300)	Sweden, Germany, UK, US (203 centres)	Persistent decline in FEV_1 exceeding 20% from baseline	✓	✓	✓	✓	✓	✓	✓	✓	✓g
SCOT (2007)[50]h	2008–ongoing	Celecoxib	Traditional NSAIDs (i.e. not COX-2 selective NSAIDs)	(~16000)	Scotland, Denmark	First occurrence of death or hospitalization for composite of non-fatal MI, non-fatal stroke, cardiovascular death	✓	✓	✓	✓	✓	✓	✓	✓	✓g
Latent TB Study (2008)[51]i	2004–7	Rifampicin (rifampin)	Isoniazid	847	Brazil, Canada, Saudi Arabia (9 hospital-based TB clinics)	Frequency of grade 3 or 4 adverse event	✓	–	✓	✓	✓	✓	✓	–	✓
Treatment Strategies for Epilepsy (2009)[52]f	2002–5	Adjust therapy	Continue therapy unchanged	111	Netherlands (7 hospitals)	Risk of complaint as measured by adverse effects scale	✓	–	✓	✓	✓	✓	✓	✓	–

PX187-0010

Table III. Contd

Study design elements						LST characteristics							Analytic method	
study (y)[a,b] year conducted	study drug	comparison	sample size: actual (intended)	participating countries (no. of sites and/or investigators)	primary endpoint(s)	open-label	general practice	broad inclusion criteria[c]	concomitant medications allowed	flexible dosing	minimal procedures required[d]	infrequent or no visits	per intent-to-treat[e]	per time on treatment
GISAS (2010)[53] f	2007–ongoing	Aripiprazole	Olanzapine, haloperidol	(800)	Italy (80 community psychiatric centres)	Onset of metabolic syndrome	✓	✓	✓	✓	✓	✓	✓	✓	NR

a Chronological order by publication year or ClinicalTrials.gov posting year, whichever is first.

b See table II for definitions of study acronyms.

c Per approved product label/monograph or minimal patient exclusion criteria.

d Limited measurements (physical or questionnaire) required beyond usual care. Detailed tests permitted when consistent with outcome assessment at baseline and/or follow-up visits.

e Includes evaluable or intent-to-treat as randomized.

f Study identified by PubMed search.

g Secondary analysis.

h Study identified by ClinicalTrials.gov search.

COX = cyclooxygenase; FEV_1 = forced expiratory volume in 1 second; **GI** = gastrointestinal; **GPs** = general practitioners; **HBVAXPRO** = hepatitis B vaccine; **HEXAVAC** = hexavalent vaccine for diphtheria, tetanus, acellular pertussis, inactivated poliomyelitis, hepatitis B and haemophilus influenza type b conjugate vaccine; **MI** = myocardial infarction; **NA** = not applicable; **NR** = not reported; **OTC** = over-the-counter; **PENTACOQ** = combination vaccine for diphtheria, tetanus, acellular pertussis, inactivated poliomyelitis and haemophilus influenza type b; **TB** = tuberculosis; **TD** = tardive dyskinesia; ✓ indicated met criteria; – indicates did not meet criteria.

follow-up and the percentage of participants' loss to follow-up are also reported in table IV.

Table IV indicates there are similarities across the studies, particularly in the reasons cited for conducting the study. The objective of each study was to address clinical uncertainty about the safe use of a health intervention. Six studies were post-approval commitments to the FDA or an EU regulatory agency to study the real-world safety of a medicine or vaccine after its approval and launch onto the market or following the emergence of safety concerns postmarketing.[43,46-50] In every study, the authors noted there was insufficient or no data available on the real-world safety of the interventions they were studying. Many investigators also reported that reliable estimates of the background incidence of the endpoint(s) were unavailable. When estimates of incidence were reported, they were typically inferred from phase III trials or spontaneous reporting rates for medicines in the same class[39,40,43,46] but in a couple of cases were based on real-world information because the medicine was already marketed.[41,50] Ten[39-44,46-49,51,52] of the 11 completed studies noted power calculations for the study, reporting the event rate or sample size needed to detect a statistically significant difference between treatment arms. Eleven studies, based on the available information, were designed as superiority trials.[39-47,50-53] Two studies, VOLUME and SMART, used a non-inferiority design.[48,49]

All studies addressed an important clinical and public health question. Every completed study provided results that meaningfully addressed the investigators' original objectives and were relevant to clinical decision making. Even VOLUME and SMART, which were stopped early and were under-powered, provided relevant descriptive and comparative data, respectively.[48,49] In the majority of studies, the absence of an increase in safety events associated with the use of the study intervention provided support for the safe use of a new medication or treatment strategy as an equivalent therapeutic option.[39-44,46,47,51] In the TB Study,[51] the findings led to a considered re-evaluation of the safety of the standard of care for treatment of latent TB and a call for further research on the comparative efficacy

810

Reynolds et al.

Table IV. Rationale, results and conclusions of the completed large simple trials

Study (y)[a]	Background and rationale	Objective	Outcome data collection method	Follow-up duration	Loss to follow-up (%)[b]	Study results and conclusions
HYPREN (1991)[39]	Case of hypotension reported after adjunctive treatment of CHF patients with ACE inhibitors Regulatory agencies recommend that treatment be initiated in hospital only No large, well controlled trials of safety Incidence of hypotension among those treated with enalapril ranged from 7% to 44% in efficacy RCTs	Compare incidence of symptomatic hypotension for enalapril to prazosin	Physician assessment	1 d (2 wk optional follow-up thereafter)	0	0.5% (95% CI 0.1, 1.5%) of enalapril exposed vs 2.6% (95% CI 1.5, 4.2%) of prazosin exposed experienced severe hypotension on d 1, and 4.7% (95% CI 3.2, 6.7%) and 10.3% (95% CI 8.0, 13.1%), respectively, experienced moderate hypotension Statistically less symptomatic hypotension observed in patients treated with enalapril than among patients treated with prazosin Risk of hypotension associated with the ACE inhibitor enalapril is lower than a widely used α-blocker
Pediatric Ibuprofen Study (1995)[40]	Ibuprofen approved by US FDA to reduce fever in children; however, because of limited safety data, only available through prescription. The other widely used NSAID for fever in children, paracetamol (acetaminophen), is available OTC In adults, NSAIDs reported to be associated with rare but serious reactions, including acute GI bleeding, acute renal failure and anaphylaxis. Randomized study proposed because of concern that an observational study would be confounded by severity, since paediatric ibuprofen is available only by prescription	Investigate if rates of hospitalization for acute GI bleeding, acute renal failure, anaphylaxis, or Reye's syndrome among patients treated with ibuprofen are elevated when compared with those treated with paracetamol	Parent/guardian report by telephone interview or questionnaire; physician report; hospital discharge summary; medical records when available	1 mo	<1	Zero deaths during 2½-y study. 1% were admitted to a hospital in the mo following enrolment; for children in the paracetamol group, the rate was 1% and for the ibuprofen 5 mg and 10 mg groups the rates were 0.9% and 1%, respectively No evidence of a significant increased risk for the four cause-specific primary outcomes Short-term use of ibuprofen not associated with increased risk of hospitalization for four serious outcomes when compared with commonly used OTC fever reducer

Continued next page

PX187-0012

Large Simple Trials in Post-Approval Safety Research   811

Table IV. Contd

Study (y)[a]	Background and rationale	Objective	Outcome data collection method	Follow-up duration	Loss to follow-up (%)[b]	Study results and conclusions
PAIN (1999)[41]	NSAIDs widely used to provide relief for patients with arthritis. They are associated with a high incidence of GI side effects In some studies, risk of GI events was found to be higher among aspirin (acetylsalicylic acid) users and lower among ibuprofen users when compared with paracetamol	Compare proportion of patients with at least one significant adverse event among those randomized to aspirin, ibuprofen and paracetamol	Patient report in a diary card	1 wk	<1	Rates of significant adverse events were aspirin 18.7%, ibuprofen 13.7% and paracetamol 14.5%. Ibuprofen was statistically equivalent to paracetamol. Both were significantly better tolerated than aspirin. Total GI events and abdominal pain were less frequent with ibuprofen (4% and 2.8%, respectively) than with paracetamol (5.3% and 3.9%) or aspirin (7.1% and 6.8%) Aspirin is not a first-line therapeutic option when the risk of GI events is of concern
ZODIAC (2002)[42-44]	Ziprasidone, used to treat schizophrenia, prolongs the QTc interval moderately when compared with other antipsychotics Significant QTc prolongation has the potential to induce ventricular arrhythmias and sudden death but the relationship between moderate QTc prolongation and risk of death or serious CV events is unknown Post-approval commitment to Swedish MPA and US FDA	Determine if ziprasidone is associated with a higher rate of non-suicide mortality within 12 mo when compared with olanzapine	Physician report; hospital discharge summaries; medical records and National Death Indices when available	1 y	<10[c]	The incidence of non-suicide mortality within 1 y of initiating pharmacotherapy was 0.91 for ziprasidone and 0.90 for olanzapine. The relative risk was 1.02 (95% CI 0.76, 1.39). This finding was confirmed in numerous secondary and sensitivity analyses Despite the known risk of QTc prolongation with ziprasidone treatment, the study found that ziprasidone was not associated with an elevated risk of non-suicidal mortality relative to olanzapine in real-world use
Olanzapine and Tardive Dyskinesia Study (2004)[45]	TD involves involuntary, often rapid and abnormal movements of the body, particularly the mouth and tongue Annual risk of developing treatment-emergent TD	Determine if TD more common among those treated with typical antipsychotics than those treated with olanzapine	Psychiatrist or psychologist assessment	9 mo	14[c]	The presence of TD was 2.3% for olanzapine and 16.7% for the typical treatment group Olanzapine is less likely to be associated with TD

Continued next page

PX187-0013

Table IV. Contd

Study (y)[a]	Background and rationale	Objective	Outcome data collection method	Follow-up duration	Loss to follow-up (%)[b]	Study results and conclusions
	among users of typical treatments reported to be between 4% and 8% Studies of the new atypical antipsychotic olanzapine reported that TD was less common, with an incidence from 1% to 2.6% when compared with typical antipsychotics					
COSMIC (2005)[46]	Lactic acidosis, a rare, serious condition that results in up to half of affected patients dying is believed to be a biguanide class effect Low (3/100 000 person-years) to zero reports of lactic acidosis associated with metformin in observational studies and a meta-analysis randomized trial and cohort studies A post-approval commitment to the FDA	Compare incidence of lactic acidosis or other serious adverse events of diabetic patients treated with metformin or usual care	Patient report through in-person or telephone interviews	1 y	<5	Zero cases of lactic acidosis reported SAEs were similar between groups, with 10.3% (95% CI 9.6, 11.1%) among the metformin group and 11.0% (95% CI 9.5, 12.7%) among the usual-care group. All-cause mortality and hospitalization were also similar Metformin was comparable with other treatments for type 2 diabetes mellitus
Pediatric Vaccination Strategies Study (2006)[47]	Multivalent vaccines are preferable because they simplify immunization schedules and increase coverage Increased coverage leads to a reduction in morbidity due to vaccine preventable diseases A post-approval commitment to the EMA	Incidence of fever of 40°C or higher (rectal) with single injection of hexavalent vaccine (HEXAVAC) vs injections with pentavalent vaccine (PENTACOQ) and Hepatitis B vaccine (HBVAXPRO)	Parent/guardian report using home thermometer	3 d	10	During the combined 3-d periods following the vaccine injections, the incidence of fever of 40°C or higher with HEXAVAC was not greater than with PENTACOQ and HBVAXPRO Single-injection vaccine reactogenicity is similar to that when using two vaccines
SMART (2006)[48]	Observational research, beginning in the late	Combined occurrence of respiratory-related	Patient report through telephone	7 mo (28 wk)	<25	Study terminated because of findings from interim analysis and slow enrolment

Continued next page

PX187-0014

Large Simple Trials in Post-Approval Safety Research

Table IV. Contd

Study (y)[a]	Background and rationale	Objective	Outcome data collection method	Follow-up duration	Loss to follow-up (%)[b]	Study results and conclusions
	1960s, found that use of inhaled β_2-agonists (e.g. isoproterenol, fenoterol, albuterol) is associated with an increase in asthma-related deaths Uncertainty whether LABAs are associated with increased asthma mortality	deaths and respiratory-related, life-threatening experiences (defined as intubation and mechanical ventilation) treated with salmeterol or placebo	interviews; death certificates; hospital records when available			No statistical difference in primary combined outcome; however, small, significant increase in respiratory-related deaths (RR 2.16 [95% CI 1.06, 4.41]) and asthma-related deaths (RR 4.37 [95% CI 1.25, 15.34]) Subgroup analyses indicated imbalance due to increase of these events in African Americans exposed to salmeterol. This subgroup also has greater disease severity at baseline and more asthma-related ED visits Findings suggest that salmeterol or other LABAs should be used only as an additive therapy to inhaled corticosteroids, when these are not sufficient for asthma control
VOLUME (2006)[49]	Inhaled insulin associated with small, reversible declines in pulmonary function in the phase III programme Given the novel route of insulin administration (via inhalation), the study was designed to evaluate longer-term pulmonary safety in real-world clinical practice A post-approval commitment to the EMA and FDA	Persistent decline in FEV_1 exceeding 20% from baseline	Physician report	5 y intended at study start, study ended early at about 15 mo	<5[d]	The study had insufficient power to address the primary endpoint in the absence of any unanticipated large effects and no formal hypothesis testing could be performed The number of subjects who experienced a persistent decline in FEV_1 exceeding 20% from baseline was higher in the inhaled insulin group compared with subjects in the comparator group using the original protocol definition of the primary endpoint. Overall, there were no clinically meaningful differences between treatment groups
Latent TB Study (2008)[51]	Standard of care for LTBI is 9 mo of isoniazid therapy. Completion rates are about 50%, however, and the therapy has been associated with drug-induced hepatitis There is a need for shorter and safer	Compare the frequency of grade 3 or 4 adverse events of two drugs with different treatment durations (4 mo of rifampicin vs 9 mo of isoniazid)	Physician report	4 or 9 mo depending on treatment	~5[c]	Study stopped about 3 y after enrolment of the first patient because of DSMB's review of a planned interim analysis Risk difference –2.3% (95% CI –5%, –0.1%; p = 0.040) for grade 3 or 4 adverse events when comparing isoniazid with rifampicin. This difference was attributable to the significant increase in drug-induced hepatitis in the isoniazed-treated group

Continued next page

PX187-0015

814   *Reynolds et al.*

Table IV. Contd

Study (y)[a]	Background and rationale	Objective	Outcome data collection method	Follow-up duration	Loss to follow-up (%)[b]	Study results and conclusions
	regimens; the only plausible alternative is 4 mo of daily rifampicin (rifampin) Published literature on rifampicin is limited: one trial found rifampicin similar to isoniazid in efficacy and safety, and two observational studies found that rifampicin is associated with better completion and less hepatotoxicity under routine care conditions					Results suggest that treatment with rifampicin may be less hepatoxic than standard isoniazid therapy Authors call for a large-scale study to compare the efficacy of rifampicin with that of isoniazid
Treatment Strategies for Epilepsy (2009)[52]	Side effects due to AEDs are common Adjustment of AEDs may reduce side effects but may also lead to recurrence of seizures Determine if well managed patients experiencing moderate to severe complaints have a decrease in side effects when current therapy dose is reduced and/or their treatment is switched	Compare reduction in risk of complaint as measured by side effects scale when employing two different treatment strategies (change dose/switch therapy vs maintain current therapy)	Patient report	7 mo	<5	After 7 mo, the RR of decrease in complaints was 1.34 (95% CI 0.88, 2.05) for the intervention group Authors concluded that despite possible risk of seizure recurrence in seizure-controlled patients, an adjustment of therapy leads to better patient-reported outcomes

a   See table II for definitions of study acronyms.

b   Patients who removed informed consent, who were unable to be contacted, including alternate contact(s), or for who no final outcome data (i.e. National Death Index) are available. Loss to follow-up estimate does not include patients who have discontinued study medication, as long as they continue to be followed.

c   Calculated using figures reported in published papers.

d   Loss to follow-up was <5% during the course of the study. After the announcement of the withdrawal of inhaled insulin a substantial number of patients immediately decided to no longer participate and withdrew from the study. At the time the study closed, loss to follow-up was 26%.

AEDs = antiepileptic drugs; **CHF** = chronic heart failure; **CV** = cardiovascular; **DSMB** = Data Safety Monitoring Board; **ED** = emergency department; **EMA** = European Medicines Agency; **FEV₁** = forced expiratory volume in 1 second; **GI** = gastrointestinal; **HBVAXPRO** = hepatitis B vaccine; **HEXAVAC** = hexavalent vaccine for diphtheria, tetanus, acellular pertussis, inactivated poliomyelitis, hepatitis B and haemophilus influenza type b conjugate vaccine; **LABAs** = long-acting third- generation β₂-agonists; **LTBI** = latent tuberculosis infection; **MPA** = Medical Products Agency; **OTC** = over-the-counter; **PENTACOQ** = combination vaccine for diphtheria, tetanus, acellular pertussis, inactivated poliomyelitis and haemophilus influenza type b; **QTc** = corrected QT interval; **RCTs** = randomized controlled trials; **RR** = relative risk; **SAEs** = serious adverse events; **TD** = tardive dyskinesia.

PX187-0016

and safety of the first- and second-line treatments.[60]

The studies varied in two notable ways: the duration of follow-up and the percentage of patients lost to follow-up. In the four short-duration studies,[39-41,47] where follow-up ranged from 1 day to 1 month, loss to follow-up was minimal. Longer duration studies had greater attrition.[42-46,48-51] Patients with schizophrenia and asthma were the most difficult groups to follow.[43,48] In studies where patients were randomized to a medicine and were followed for longer than 1 month, we found a range in adherence to the treatment under study versus the comparator: 90% versus 77% in COSMIC,[46] 64% versus 73% in ZODIAC,[43] 89% versus 95% in the Tardive Dyskinesia Study;[45] 73% versus 72% in SMART;[48] and 78% versus 60% in the TB Study.[50] Concern about the bias introduced by discontinuation of study medications, and thus the validity of ITT analyses, may explain why person-time analyses, where rates for on-treatment periods are calculated and compared, were performed in many studies. For example, despite having higher than expected adherence rates for the patient populations being studied, both ZODIAC and the TB studies reported results from person-time analyses.[43,50] These on-treatment analyses are also potentially biased, however. They are subject to the same biases as a non-randomized study and, thus, require adjustment for potential confounders.

Although most studies successfully addressed their research objectives, challenges complicating the interpretation of the findings from safety LSTs were noted, including low event rates, heterogeneous effects within the study population, lengthy enrolment periods and study durations, and the use of concomitant medications. Despite the large sample size of many of the LSTs, the absolute number of events reported in studies was often not very large, particularly if death was the outcome. For example, in the Pediatric Ibuprofen Study, despite its large size, the investigators note they could not exclude serious events occurring at a rate <1/10 000.[40] The COSMIC Study did not observe any cases of lactic acidosis, which was its primary regulatory purpose.[46]

In ZODIAC,[43] even a 'hard' outcome such as sudden death was still infrequent enough that the study was designed and powered to detect differences in non-suicide mortality, despite being conducted in a patient population known to have a higher rate of cardiovascular events than the general population.[61] A lower than expected event rate in the SMART trial led to a re-calculation of its required sample size to 60 000 patients, double the initially intended size. Furthermore, the SMART trial's finding that African Americans were at a higher risk of respiratory and asthma-related events, likely due to genetic or behavioural factors, highlights the problems of conducting a LST where treatment effects vary within or across the study population.[48] When an investigator expects subgroup differences, the LST design is likely not the preferred study option.[24,26]

The length of the study, particularly as it relates to the evolving benefit-risk balance of the medicines under study, and allowance of concomitant medication use may complicate interpretation of study results. During the course of the ZODIAC Study, concerns were raised about the potential for antipsychotics to be associated with the development of diabetes and metabolic syndrome. Although this eventually resulted in class labelling for atypical antipsychotics, physician perception that there were differential risks among antipsychotics was anecdotally cited as a reason for decreased enrolment at investigator sites. Finally, in the Pediatric Vaccination Study,[47] the authors noted that parents in France commonly use prophylactic antipyretics to prevent fever after vaccination. Since the study was meant to be reflective of real-world practice, the investigators did not prevent their use. Instead, they tried to distinguish between curative and preventative use of fever reducers in *post hoc* analyses. This challenge highlights the difficulty of estimating the incidence of outcomes if the definition of the endpoint (in this case fever) is preventable through other interventions used in real-world practice. However, from a pragmatic perspective this is acceptable, particularly if it means that the intervention as it is actually used, with other medicines, is safe.

7. Conclusions

Very few completed or ongoing LSTs evaluating safety outcomes as a primary endpoint were identified in our review. Among those that have been conducted, we found that their design elements were similar over time. Three LSTs evaluated the safety of prescription or non-prescription NSAIDs. This is likely due to regulatory and public health concerns about their safety given their widespread use. Although the LST design may be most appropriate when studying 'hard' outcomes such as death or hospitalization, we found that researchers also used the design to compare 'soft' outcomes (e.g. incidence of physician reported symptomatic hypotension and patient reported complaints) or outcomes that require regular measurement (e.g. pulmonary function). This finding was unexpected since the seminal literature on the use of LSTs for examining the therapeutic or preventative effects of an intervention suggest they are best suited to studying 'hard' outcomes to avoid assessment and reporting bias.[24-26] Other unexpected differences we noted were the use of person-time on treatment secondary analyses, which result in potentially biased comparisons between treatment groups (i.e. because participants discontinue their medications for reasons that may be related to potential outcomes), and sample sizes that were smaller than thousands of patients.

The results of the search demonstrate that there is significant variation in the terms used to describe LSTs or denote the intent to minimize interference with routine care. Despite potential differences in nomenclature, we hypothesized that the intent of 'simple' or 'practical' trials would be the same: to minimize random error by using a large sample size, to distribute known and unknown confounders across baseline treatment groups by randomization, and to preserve routine care and maximize generalizability of findings by following patients using observational methods. This was found to be true, although the earlier studies indicated a tendency toward the use of elements associated with a controlled clinical trial (e.g. double-blinds) whereas more recent LSTs had the most characteristics associated with observational epidemiology. Nonetheless, consistent with our selection criteria, all LSTs included in this review utilized observational methods of follow-up, but did so with varying degrees of intervention to meet the study objectives, primarily in the form of scheduled visits or required laboratory and diagnostic tests. For example, to permit analytic comparison across treatment groups, the VOLUME Study required more intervention (e.g. requirement for spirometry among inhaled insulin and usual diabetes care users to make valid comparisons of pulmonary safety) despite its interference with usual care (i.e. usual diabetes care does not routinely include pulmonary function tests).[49] ZODIAC, on the other hand, evaluated only mortality and hospitalization rates, and the study protocol did not mandate interventions post-randomization.[42,43]

A good LST has been defined as one that asks an important public health question and does so reliably.[24,26] Comparison of the LSTs' original objectives with their results indicates that the design yields important clinical insights. These studies' findings have provided data critical to evidence-based comparisons of treatments as used in the real-world and, in some cases, led to the re-evaluation of medical practice.[50,60] Despite the success of these studies, it is unreasonable to expect the findings of an LST to always result in changes to guidelines on clinical practice or policy decisions. Rigorous safety evaluation typically requires critical interpretation of results from multiple sources, and the results of a single safety LST may not be sufficient to immediately alter medical practice.

Finally, while medication adherence in these studies was relatively high it was different by treatment arm in some studies. ITT analyses are most valid when there is maximum, and equal, adherence to the assigned study medication across treatment arms. Bias as a result of non-adherence is a justifiable concern for any randomized study, including LSTs. When appreciable non-adherence occurs, and is non-differential, the true effect will be underestimated by ITT. In a safety LST this would result in a bias toward the null, decreasing the likelihood of observing an adverse effect of a medicine. Furthermore, if non-adherence is

differential, then despite randomization, both the ITT and person-time analyses will be biased, requiring adjustment for confounders. In these analyses, because patients choose to discontinue or switch medicines for unknown reasons, the benefits of randomization are lost and the groups being compared can no longer be assumed to have an equal distribution of known and unknown confounders. Research on how to address non-compliance in randomized studies with observational follow-up is growing; for example, one recently published study has described the use of inverse probability weighting to adjust for incomplete adherence to assigned treatment.[62]

The finding that LSTs are rarely used to evaluate safety endpoints may be due to a lack of experience with the design among drug safety professionals but it is more likely a result of the operational, financial and scientific hurdles of implementing the design. Substantial resources are required to accrue large sample sizes, collect multiple forms of outcome data, manage hundreds of participating investigators and sites, and ensure appropriate scientific and ethical oversight. The lack of research infrastructure for conducting research at sites or with physicians inexperienced with randomized trials, including the complex regulations governing interventional study implementation and prohibitive financial costs, have been described previously as barriers to conducting LSTs studying efficacy.[63,64] This complexity suggests that the design might be more accurately described as a 'simplified' rather than a 'simple' trial. Efforts have been made to simplify LSTs by recruiting patients using electronic healthcare databases, as is being done in SCOT with the Scottish Medicines Monitoring Unit (MEMO) database,[50] but others have found that patient recruitment in this way was not practical.[65] Financial considerations are also important. Twelve of the LSTs identified in this review were funded by the pharmaceutical or consumer products industry,[39-50,52,53] six of which were identified as post-approval commitments to the FDA or an EU regulatory agency.[43,46-50] Whether there is a lack of interest in funding from public research institutes or a societal expectation that costs of researching the safety of health in-terventions should be borne by private industry is unclear.

Lesko and Mitchell,[40] in their overview of lessons learned from the Pediatric Ibuprofen Study, point out that a simple testable safety hypothesis, a motivated patient and physician population, and the ability to follow-up patients to assess outcomes are critical to the feasibility of LSTs for safety.[66] Many other factors influence whether a LST is feasible. For example, evaluating differences in very rare (e.g. Steven's Johnson Syndrome) or long latency (e.g. cancer induction) outcomes are not practical with the LST design. Even hard outcomes such as death are difficult to study in a randomized study if the incidence of the outcome is low or the observed rate in the study is lower than expected when initial sample size calculations are performed. Because the design uses randomization, an appropriate comparator acceptable to physicians and patients is also a critical factor in the success of a LST. Equipoise must exist, and if a new medication or vaccine has a real or perceived benefit over available treatments, the LST design will not be feasible. The lack of equipoise may also become a challenge in situations where the study enrolment period is very lengthy. As knowledge of the benefits and risks of treatments evolve, leading to less clinical uncertainty about the type of patients that might benefit from a particular study medication, investigators may find it increasingly difficult to enrol patients.

The scientific question underlying a safety LST must be relevant clinically and of interest to physicians practicing in private and community health settings. It is inclusion of those sites that make the findings generalizable. Interestingly, studies that seek to compare the general safety profile of medicines or vaccines, in lieu of a specific research hypothesis for which the study is powered, do seem to be feasible. Three studies[54,55,57] excluded from our review because they collected and compared all serious adverse events rather than a pre-defined safety endpoint, were able to successfully enrol many patients. Based on our review, there is insufficient evidence to restrict the design to studies with short-term follow-up or to patient populations with lower rates of therapy

discontinuation. Finally, data collection directly from patients, rather than physicians, is not necessarily a barrier either, as the PAIN Study demonstrated.[67,68]

While the LST is potentially a design to study the safety of any health intervention, its use with novel medicines, vaccines, diagnostics and biologics is likely to be limited. The feature of randomization alone limits its application, since in some disease areas there may be no logical or ethical comparator (i.e. first-in-class medicine or vaccine, rare diseases, oncology). Biopharmaceutical development is increasingly focused on smaller indicated patient populations. This will result in insufficient exposure to reach the required event rate or, when feasible, it may take too long to accrue exposure to be acceptable to decision makers. Even in situations where the prevalence of a condition is much greater (e.g. schizophrenia or asthma) it may take many years to accrue the necessary patient population, which to some may be too long to address important safety questions.[69] Regardless of how the design is used in the future, well designed observational studies, whether they involve primary data collection or use secondary data sources such as large, electronic healthcare databases, will continue to be an important and, in some cases, the only resource to investigate many post-approval safety questions.

Despite the challenges of using the LST design for comparative safety evaluation, it is particularly suited to research questions in which confounding by indication or severity is likely to be pronounced and difficult to measure or control, but where assessment under routine care is important for decision making. LST designs are similar to observational studies in that they can, in principle, be effectively used to study the safety of health interventions in patient populations not typically exposed in clinical trials, such as the elderly, very young or those with multiple co-morbidities; determine if physicians prescribe according to their interpretation of the product label or clinical experience; and understand the safety of a health intervention as it is used with multiple concomitant prescriptions or over-the-counter medications under routine medical care.

With the increasing demand for real-world evidence to guide public health and clinical decisions, the design's advantages and disadvantages for future comparative safety research should be carefully considered by researchers.

Acknowledgements

References

1. Tufts Center for the Study of Drug Development. Tufts impact report. Boston (MA): Tufts Center for the Study of Drug Development, 2008 Jul/Aug

2. US Food and Drug Administration. Report to congress: report on postmarketing studies. Washington, DC: US Food and Drug Administration, 2002

3. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2004; 58: 323-37

4. Sobel RE, Reynolds RF. Integrating evidence from multiple sources to evaluate post-approval safety: an example of sildenafil citrate and cardiovascular events. Curr Med Res Opin 2008; 24 (7): 1861-8

5. Reynolds RF, Glasser DB, Dieck GS. A view from industry. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. West Sussex: John Wiley & Sons Ltd, 2005: 77-101

6. Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. JAMA 1999 Mar 3; 281 (9): 824-9

7. Kaitin KI. The new drug approvals of 1999, 2000, and 2001: drug development trends a decade after passage of the Prescription Drug User Fee Act of 1992. Drug Inf J 2003; 37: 357-71

8. Csizmadi I, Collet JP, Boivin JF. Bias and confounding in pharmacoepidemiology. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. West Sussex: John Wiley & Sons Ltd, 2005: 791-809

9. D'Agostino Jr RB, D'Agostino Sr RB. Estimating treatment effects using observational data. JAMA 2007; 297 (3): 314-6

10. Walker A. Confounding by indication. Epidemiology 1996; 7: 335-6

11. Joffe MM. Confounding by indication: the case of calcium channel blockers. Pharmacoepidemiol Drug Saf 2000; 9 (1): 37-41

12. Kizer JR, Kimmel SE. Epidemiologic review of the calcium channel blocker drugs. Arch Intern Med 2001; 161: 1145-58

13. Ahdieh L, Gange SJ, Greenblatt R, et al. Selection by indication of potent antiretroviral therapy use in a large cohort of women infected with human immunodeficiency virus. Am J Epidemiol 2000; 152 (10): 923-33

14. Garrett J, Lanes S, Kolbe J. Risk of severe life threatening asthma and beta agonist type: an example of confounding by severity. Thorax 1996; 51: 1093-9

15. Dunn N, White I, Freemantle S, et al. The role of prescribing and referral bias in studies of the association between third generation oral contraceptives and increased risk of thromboembolism. Pharmacoepidemiol Drug Saf 1998; 7: 3-14

16. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of β_2-agonists. Am J Epidemiol 1996; 144 (12): 1161-9

17. LaPorte J, Ibanez L, Vidal X, et al. Upper gastrointestinal bleeding associated with the use of NSAIDs: newer versus older agents. Drug Saf 2004; 27: 411-20

18. Mamdani M, Rochon P, Juurlink D, et al. Observational study of upper gastrointestinal haemorrhage in elderly patients given selective cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs. BMJ 2002; 325: 624-9

19. Vandenbroucke JP, Helmerhorst FM, Bloemenkamp KWM, et al. Third-generation oral contraceptive and deep venous thrombosis: from epidemiologic controversy to new insight in coagulation. Am J Obstet Gynecol 1997; 177: 887-91

20. Walker AM. Newer oral contraceptives and the risk of venous thromboembolism. Contraception 1998; 57: 169-81

21. Wolfe F, Flowers N, Burke T, et al. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002; 29: 1015-22

22. Avorn J. In defense of pharmacoepidemiology: embracing the ying and yang of drug research. N Engl J Med 2007; 357 (22): 2219-21

23. Madre LK, Califf RM, Reynolds RF, et al. Views from academia, industry, and regulatory agencies. In: Strom BL, Kimmel SE, editors. Textbook of pharmacoepidemiology. West Sussex: John Wiley & Sons, Ltd, 2006: 63-88

24. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984; 3: 409-20

25. Buring JE, Hennekens CH. The contribution of large, simple trials to prevention research. Prev Med 1994; 23: 595-8

26. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. J Clin Epidemiol 1995; 48 (1): 23-40

27. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2005; 290 (12): 1624-32

28. ISIS-1 Group. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. Lancet 1986; II (8498): 57-66

29. ISIS-2 Group. Randomised trial of intravenous streptokinase, oral, aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988; II (8607): 349-60

30. ISIS-3 Group. A ramdomised comparison streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. Lancet 1992; 339: 753-70

31. ISIS-4 Group. A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. Lancet 1995; 345: 669-82

32. Steering Committee of the Physicians Health Study Research Group. Final report on the aspirin component of the ongoing physicians' heath study. N Engl J Med 1989; 321 (3): 129-35

33. Smart Study Group. CD4+ count-guided interruption of antiretroviral treatment. N Engl J Med 2006; 355 (22): 2283-96

34. Hasford J. Drug risk assessment: a case for large trials with lean protocols. Pharmacoepidemiol Drug Saf 1994; 3: 321-7

35. Mitchell AA, Lesko SM. When a randomized controlled trial is needed to assess drug safety: the case of paediatric ibuprofen. Drug Saf 1995; 13: 15-24

36. US Food and Drug Administration. Guidance for industry: premarketing risk assessment. Washington, DC: US Food and Drug Administration, 2005

37. Committee on the Assessment of the US Drug Safety System, Board on Population Health and Public Health Practice. Baciu A, Stratton K, Burke SP, editors. The future of drug safety. Washington, DC: The National Academies Press, 2007

38. Gillen JE, Tse T, Ide NC, et al. Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. Stud Health Technol Inform 2004; 107: 1466-70

39. Hasford J, Bussman WD, Delius W, et al. First dose hypotension with enalapril and prazosin in congestive heart failure. Int J Cardiol 1991; 31: 287-94

40. Lesko SM, Mitchell AA. An assessment of the safety of pediatric ibuprofen, a practitioner-based randomized clinical trial. JAMA 1995; 273 (12): 929-33

41. Moore N, Vanganse E, Le Parc J-M, et al. The PAIN study: paracetamol, aspirin and ibuprofen new tolerability study: a large-scale, randomised clinical trial comparing the tolerability of aspirin, ibuprofen and paracetamol for short-term analgesia. Clin Drug Invest 1999; 18 (2): 89-98

42. Strom BL, Faich GA, Reynolds RF, et al. The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC): design and baseline characteristics. J Clin Psychiatry 2008; 69: 114-21

43. Strom BL, Eng SM, Faich G, et al. Comparative mortality associated with ziprasidone and olanzapine in real-world use among 18,154 patients with schizophrenia: the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC). Am J Psychiatry 2010; 168: 193-201

44. Pfizer. Large simple trial (LST) of cardiovascular safety of ziprasidone and olanzapine (ZODIAC) [ClinicalTrials.gov identifier NCT00418171]. US National Institutes of Health, ClinicalTrials.gov [online]. Available from URL: http://www.clinicaltrials.gov [Accessed 2011 Jan]

Reynolds et al.

45. Mari JJ, Lima M, Costa AN, et al. The prevalence of tardive dyskinesia after a nine month naturalistic randomized trial comparing olanzapine with conventional treatment for schizophrenia and related disorders. Eur Arch Psychiatry Clin NeuroSci 2004; 254: 356-61

46. Cryer DR, Nicholas SP, Henry DH, et al. Comparative outcomes study of metformin intervention versus conventional approach: the COSMIC Approach Study. Diabetes Care 2005; 28 (3): 539-43

47. Reinert P, Fiquet A, Thomas S, et al., on behalf of the HEXALIS Study Group. Fever as a marker of reactogenicity of an acellular pertussis-containing hexavalent vaccine (HEXAVAC) in a large-scale, open, randomized safety study in healthy French infants. Hum Vaccin 2006; 5: 215-21

48. Nelson HS, Weiss ST, Bleecker ER, et al. The salmeterol multicenter asthma research trial: a comparison of usual pharmacotherapy for asthma or usual pharmacotherapy plus salmeterol. Chest 2006; 129: 15-26

49. Pfizer. Exubera large simple trial to evaluate long-term pulmonary and cardiovascular safety (VOLUME) [Clinical Trials.gov identifier NCT00359801]. US National Institutes of Health, ClinicalTrials.gov [online]. Available from URL: http://www.clinicaltrials.gov [Accessed 2011 Jan]

50. University of Dundee. The standard care versus celecoxib outcome trial (SCOTLSSS) [ClinicalTrials.gov identifier NCT00447759]. US National Institutes of Health, ClinicalTrials.gov [online]. Available from URL: http://www.clinicaltrials.gov [Accessed 2011 Jan]

51. Menzies D, Long R, Trajman A, et al. Adverse events with 4 months of rifampin therapy or 9 months of isoniazid therapy for latent tuberculosis infection. Ann Intern Med 2008; 149: 689-97

52. Uijl SG, Uiterwaal CSPM, Aldenkamp AP, et al. Adjustment of treatment increases quality of life in patients with epilepsy: a randomized controlled pragmatic trial. Eur J Neurol 2009; 16: 1173-7

53. Mario Negri Institute for Pharmacological Research. Pragmatic RCT comparing aripiprazole, olanzapine and haloperidol in the treatment of schizophrenia (GiSAS) [Clinical Trials.gov identifier NCT01052389]. US National Institutes of Health, ClinicalTrials.gov [online]. Available from URL: http://www.clinicaltrials.gov [Accessed 2011 Jan]

54. Hawkey C, Kahan A, Steinbruck K, et al. Gastrointestinal tolerability of meloxicam compared to diclofenac in osteoarthritis patients. Br J Rheumatol 1998; 37: 937-45

55. Dequeker J, Hawkey C, Kahan A, et al. Improvement in gastrointestinal tolerability of the selective cyclooxygenase (COX)-2 inhibitor, meloxicam, compared with piroxicam: results of the Safety and Efficacy Large-Scale Evaluation of Cox-Inhibiting Therapies (SELECT) trial in osteoarthritis. Br J Rheumatol 1998; 37: 946-51

56. Lauterbach M, Martins IP, Castro-Caldas A, et al. Neurological outcomes in children with and without amalgam-related mercury exposure: seven years of longitudinal observations in a randomized trial. J Am Dent Assoc 2008; 139: 138-45

57. Castle W, Fuller R, Hall J, et al. Serevent nationwide surveillance study: comparison of salmeterol with salbutamol in asthmatic patients who require regular bronchodilator treatment. BMJ 1993; 306: 1034-7

58. Faich GA, Morganroth J, Whitehouse AB, et al. Clinical experience with moxifloxacin in patients with respiratory tract infections. Ann Pharmacother 2004; 38: 749-54

59. Ashraf E, Ford L, Geetha R, et al. Safety profile of ibuprofen suspension in young children. Inflammopharmacology 1999; 7: 219-25

60. Blumberg HM. Needed: new and better tools to combat latent tuberculosis infection. Ann Intern Med 2008; 149: 761-3

61. Brown S. Excess mortality of schizophrenia: a meta-analysis. Br J Psychiatry 1997; 171: 502-8

62. Toh S, Hernandez-Diaz S, Logan R, et al. Estimating absolute risks in the presence of non-adherence: an application to a follow-up study with baseline randomization. Epidemiology 2010; 21: 528-39

63. Yusuf S, Mehta SR, Diaz R, et al. Challenges in the conduct of large simple trials of important generic questions in resource-poor settings. Am Heart J 2004; 148: 1068-78

64. Egan D, Gellar N, Yusuf S, et al. Lessons learned from the DIG trial. Control Clin Trials 2003; 24: 316-26S

65. Mosis G, Dieleman JP, Stricker BC, et al. A randomized database study in general practice yielded quality data but patient recruitment in routine consultation was not practical. J Clin Epidemiol 2006; 59: 497-502

66. Lesko SM, Allen M. The use of randomized controlled trials for pharmacoepidemiology studies. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. West Sussex: John Wiley & Sons Ltd, 2005: 599-610

67. Moore N, Charlesworth A, Van Ganse E, et al. Risk factors for adverse events in analgesic drug users: results from the PAIN study. Pharmacoepidemiol Drug Saf 2003; 12: 601-10

68. Van Ganse E, Jones JK, Moore N, et al. A large simple clinical trial prototype for assessment of OTC drug effects using patient-reported data. Pharmacoepidemiol Drug Saf 2005; 14: 249-55

69. Stroup TS. What can large simple trials do for psychiatry? Am J Psychiatry 2011; 168: 2-4

Correspondence: Dr *Robert F. Reynolds*, ScD, Department of Epidemiology, Pfizer, Inc., 235 E 42nd St, MS 150/3/04, New York, NY 10017, USA.
E-mail: robert.reynolds@pfizer.com